Journal of Indonesian Economy and Business
Volume 28, Number 1, 2013, 45 – 61

SPATIAL SMALL AREA ESTIMATION FOR DETERMINATION OF UNDERDEVELOPED VILLAGES IN THE PROVINCE OF YOGYAKARTA (DIY) IN 2011[1]

Lilis Nurul Husna

Badan Pusat Statistik (BPS) RI
(lilishusna@gmail.com)

Sarpono

Head of Standardization and Statistical Classification BPS RI
(sarpono@yahoo.com)

ABSTRACT

Indonesia's poverty alleviation programs are implemented by two approaches target, those are the pockets (areas) of poverty and the poor households. Related with poverty alleviation programs targeting poor areas, in the Medium Term Development Plan (RPJM) 2010-2014, the government through the Development Backward Areas Ministry (KPDT) has determined the backward or underdeveloped regions at the level of district/city. There is no district/city in the province of Yogyakarta (DIY) are classified as underdeveloped region, but in 2011 the poverty rate in DIY is the highest compared with other provinces in Java and Bali. Therefore, the classifications of underdeveloped areas are not optimal if applicable only within the district, but it needs to be seen in the smaller scope, such as village. The main purpose of this study is to determine the underdeveloped villages in DIY in 2011. The average per capita household expenditure is a key indicator in measuring poverty. Susenas data can only be used to estimate the average per capita household expenditure to the level of district. Therefore, to obtain the estimated value in village level, this study used Small Area Estimation approach by combining census data (Podes 2011) and survey data (Susenas 2011). This study used Geographically Weighted Regression (GWR) with Adaptive Gaussian Kernel Bandwidth weighting function. GWR is a linear regression model that produces the local parameters in all locations. GWR parameters estimated are performed by Weighted Least Squares (WLS) method which involving spatial aspects. The results found that there were 13 underdeveloped villages in DIY. Furthermore, the Local Indicator of Spatial Association (LISA) is used to look at the tendency of cluster in underdeveloped villages. Then, maps are used to compare characteristic of underdeveloped villages among others.

Keywords: *poverty, underdeveloped areas, spatial, SAE, GWR*

[1] This paper is the 3rd winner of JIEB Best Paper Award 2012.

INTRODUCTION

Since early 1970s, poverty alleviation is a priority in Indonesia development strategies. The main objective of poverty reduction strategy is to improve the resident welfare and reduce socio-economic disparities or inequality intergroup residents. Poverty alleviation program directly implemented in early 1990s. At that time, the government has formulated several programs targeted using two approaches, namely pockets (areas) of poverty and the poor households. Target areas refer to to the identification of poor or underdeveloped areas. Meanwhile, individual targets aimed at households or residents who have low incomes (under the poverty line) (BPS, 2004).

Related with poverty alleviation programs which targeting poor areas, in the Medium Term Development Plan (RPJM) 2010-2014, the government through the Development Backward Areas Ministry (KPDT) has set the backward/underdeveloped regions based on six main criteria, those are: (1) economic society, (2) human resources, (3) infrastructure, (4) local financial capacity (fiscal gap), (5) accessibility, and (6) the area characteristics. In addition to these basic criteria, it is also considered that the condition of the district is located in the border areas between countries, disaster-prone areas, and areas otherwise specified.

Development of underdeveloped areas is a deliberate attempt to change a region occupied by communities with different socio-economic problems and physical limitations, to be developed by the community whose have same quality or nearly same compared with other Indonesian society. Development of underdeveloped areas program focused on accelerating development in the areas of social, cultural, economic, financial areas, accessibility, and the infrastructure supplay is still lagging behind compared to other areas.

In the 2005-2009 Development Planning, there are two districts in the Province of Yogyakarta (DIY) included in underdeveloped areas, namely Gunung Kidul and Kulon Progo. In 2009, the two of them classified as the adequate district so in 2010-2014 Development Planning, no districts in DIY is underdeveloped area. Nevertheless, based on BPS, in 2011 DIY poverty level at is still in top of the rank in the Java-Bali, amounting to 16,08 percent, higher than Jakarta (3,75 percent), Bali (4,20 percent), Banten (6,32 percent), West Java (10,65 percent), East Java (14,23 percent), and Central Java (15,76 percent).

Based on these facts, classifications of underdeveloped areas are not optimal if only applicable within the district, but it needs to be seen more in the smaller scope, such as villages. BPS has been several times determine underdeveloped villages, namely in 1993, 1994, and 2002. In 1993, BPS used 33 variables. In 1994 BPS used 17 variables to urban areas and 18 variables rural areas. In 2002 BPS set the determination of underdeloped villages uses 19 variables for urban areas and 17 variables for the rural areas.

According to BPS (2005), underdeveloped villages are villages whose condition is relatively worse from other villages. Development of an area is reflected by the main indicators, namely the level of average per capita household expenditure. The average per capita household expenditure can be obtained from the Susenas, but it can only be used to estimate in district level. Therefore, this study combines survey data (Susenas) and census data (Podes) to obtain an estimated value of village level, similar with the method developed by the World Bank in conducting *small area estimation* (Hentschel, *et al*., 2000).

Average per capita household expenditure as one of the main indicators measuring poverty is often modeled as a function of the global regression (Dimulyo, 2009). Global regression models assume that the same regression coefficients can be applied to all geographical locations. Global models will provide reliable information for smaller areas if there is no or only little variability of local regions or often referred to spatial stationarity (Fotheringham *et al*., 2002). However, the spatial stationarity condition is rare. In addition, the poverty condition of a country also affected by location of the observation or the village, including the position of the other villages around that (Rita & Anik, 2010). Thus, involving spatial factors are important in poverty analysis.

Based on that background, this study aims to determine underdeveloped villages in DIY by using local spatial regression models, namely Geographically Weighted Regression (GWR). Moveover, it also descibes the condition of underdeveloped village in DIY. This study is expected to give benefit in reduce poverty efforts in DIY and inspire governments and other researchers in the method of determining the underdeveloped villages.

DATA SOURCES

This study uses secondary data from National Socio-economic Survey (Susenas) 2011 and Villages Potential (Podes) 2011. The data was obtained from the BPS. The average per capita household expenditure got from Susenas is used as the response variable, while villages characteristics obtained from Podes used as predictor variables.

BPS (2005) determined that there are several factors suspected to be the cause of development of a village. They are natural factors/environmental, institutional factors, factors of facilities/infrastructure and access, and socio-economic factors. Based on Podes 2011 availabel data, this study used 15 variables covering those four factors to remains underdeveloped village establish model as presented in Table 1.

Table 1. Predictor variables to determine underdeveloped village

Variable	Specification
(1)	(2)
	Nature factor/environment
X_1	Population density
	Institutional factor
X_2	Government status
	Factors of facilities/infrastructure and access
X_3	Distance village office with the office of regent
X_4	Number of education facility per 100 residents
X_5	Number of medical facilityper 100 residents
X_6	Number of medical staff per 100 residents
X_7	Public telephone (wartel) existence
X_8	Number of minimarket
	Socio-economic Factors
X_9	Precentage of farmer household
X_{10}	Main income most residents
X_{11}	Percentage of family who subscribe electricity
X_{12}	Precentage of family who live along the river
X_{13}	Source of water used for drinking/cooking
X_{14}	Existence of sufferer malnutrition
X_{15}	Fuel used most of the residents

ANALYSIS METHOD

1. Global Regression Model

Global regression models that often used to examine the relationship between predictor variables and the response variable is a multiple linear regression with the model as following:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \; ;$$

$$i = 1, 2, ... n \qquad (1)$$

Which, $\beta_0, \beta_1, \ldots \beta_p$ are the parameters and $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_n$ assumed as random error distributed $N(0,\sigma^2 \boldsymbol{I})$, with I is the identity matrix. β parameter values are estimated by Ordinary Least Square method (OLS) as follows:

$$\hat{\beta} = (\hat{\beta}_0 \hat{\beta}_1 \ldots \hat{\beta}_p)^T = (X^T X)^{-1} X^T y \qquad (2)$$

Where **X** is the matrix variable predictor size *n* x *k*, *k* is the number of parameters ($k = p + 1$) and *p* is the number of predictor variables. The first column marik **X** is worth one. Meanwhile, *y* is the response variable vector. In equation (1), the value of $\hat{\beta}$ is assumed constant in all location so-called global model.

2. Geographically Weighted Regression (GWR)

Location produces two types of spatial effects, those are spatial dependence and spatial heterogeneity (Anselin, 1992). Geographically Weighted Regression (GWR) is one method used to estimate the data that has a spatial heterogeneity (Brunsdon, Fotheringham & Charlton, 1996). GWR will result local parameter estimated, where each observation will have different parameter estimated. In GWR models, each observation is georeferenced, which have coordinate points (latitude and longitude). The GWR model by Fotheringham (2002) as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j (u_i, v_i) x_{ij} + \varepsilon_i \; ;$$

$$i = 1,2,\ldots n \qquad (3)$$

Where y_i is the value of the dependent variable in the *i-th* observations, x_{ij} is the value of the *j-th* independent variable in the *i-th* observation, $\beta_0(u_i, v_i)$ is intercept on the observations at-*i*, $\beta_j(u_i, v_i)$ is parameter estimated of predictor variable x_j on *i-th* observation, *p* is the number of predictor variables, (u_i, v_i) is the coordinate points of observate location and ε_i is random error that assumed distributed $N(0, \sigma^2 I)$.

The parameters estimated by GWR models will vary at all location, so there are $n \times k$ parameters to be estimated, which *n* is the number of observation and $k=p+1$ is the number of parameters at each observation location. To estimate these parameters, GWR used Weighted Least Squares (WLS) method to give the differently weights at each observation. In providing weighting, this method follows Tobler's First Law of Geography, which location that is near to the *i-th* location will have more effect in predicting the parameters in *i* compared the further data. Thus, a nearer location will be given greater weighting.

Estimator for local regression coefficients in GWR are as follows:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \qquad \textbf{.....} \ (4)$$

Where $\hat{\beta}(u_i, v_i) = (\hat{\beta}_{i0}, \hat{\beta}_{i1}, \hat{\beta}_{i2}, \ldots, \hat{\beta}_{iP})^T$ is the local regression coefficient vector and $W(u_i, v_i)$ is a diagonal matrix with elements on that diagonals are geographical weighting on each location for observation location at-*i*, and other elements are zeros.

One of the weighting function that often used is Gaussian kernel function as follows:

$$w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right] \quad (5)$$

Where d_{ij} is the Euclidean distance between *i-th* and *j-th* location and b is the bandwidth. Weighted value will be close to 1 if the distance is near to the observation or coincide, and will become smaller so close to zero if the distance is farther away.

Determining the optimum bandwidth before forming the GWR models is very important (Fotteringham *et al*., 2002). One of methods can be used to determine the optimum bandwidth is the minimum Akaike Information Criterion (AIC). According Huvrich *et al*. (1998), AIC formula for GWR is as follows:

$$AIC_c = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + n\left\{\frac{n + tr(S)}{n - 2 - tr(S)}\right\} \quad (6)$$

Where n is the sample size, $\hat{\sigma}$ is the standard deviation, and tr (***S***) is the trace of the hat matrix.

Goodness of Fit Test needs to be done to see if the GWR models better than the OLS models. Testing is done as follows (Brunsdon, *et al*., 1999):

$$F = \frac{(SSR_{OLS} - SSS_{GWR})/v_1}{SSR_{GWR}/\delta_1} \quad (7)$$

Where SSR_{OLS} dan SSR_{GWR} row is the sum of squared residuals OLS and GWR model. Model F will approach the F distribution with freedom degrees v_1^2/v_2, δ_1^2/δ_2 where:

$$v_1 = n - p - 1 - \delta_1 \quad (8)$$

$$v_2 = n - p - 1 - 2\delta_1 + \delta_2 \quad (9)$$

$$\delta_i = tr[(I-S)^T(I-S)]^i \ , i = 1, 2. \quad (10)$$

When the significance level is α, zero hipothesis stating that the GWR and OLS models equally well in explaining the relationship between the predictor variables and the response variable will be rejected if $F > F_\alpha$ (v_1^2/v_2, δ_1^2/δ_2).

Prediction

Although the GWR main method used to explore the existence of spatial nonstationarity of the parameters, the prediction is an important aspect in the regression analysis. Prediction is done to get the value of the response variables in new areas, such as areas are not covered in the survey (Leung, et.al, 2000).

If the GWR model can explain well one set of data, then the model can be used to predict the value of the variable response to a new location by using the predictor variables in the new location. Suppose $(x_{01}x_{02}\dots x_{0p})$ is the value of the predictor variables at the new location (p_0) and $X_0^T = (1x_{01}x_{02}...x_{0p})$, then the predictive value of y_0 is:

$$\hat{y}_0 = X_0^T\hat{\beta}(p_0) \quad (11)$$

Which

$$\hat{\beta}(p_0) = (X^TW(p_0)X)^{-1}X^TW(p_0)y \quad (12)$$

And the $W(p_0)$ parameters are known and determined through model calibration process. Meanwhile, the distance between p_0 with another point in the sample is also known if p_0 coordinates are known.

3. Local Indicators of Spatial Association (LISA)

LISA is a statistical value used to test region propensity to experience the interaction region or outlier individually. LISA can be used to see the local spatial autocorrelation. Significant LISA values showed the village interacting with others significantly or an outlier.

LISA statistic values for *i-th* location, is shown as follows:

$$I_i = \frac{(x_i - \bar{x}_i)}{\frac{\Sigma_i (x_i - \bar{x}_i)^2}{n}} \sum_j w_{ij} (x_j - \bar{x}_j) \qquad (13)$$

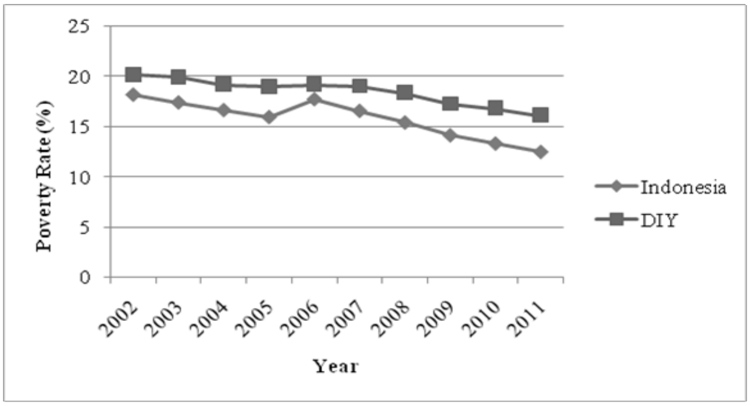
Where $\bar{x}$ is the mean of the x variable and w_{ij} is the weight matrix elements. The weights based on spatial boundaries, called the Queen Contiguity (side-angle intersection).

RESULTS

1. Poverty in the Province of Yogyakarta (DIY)

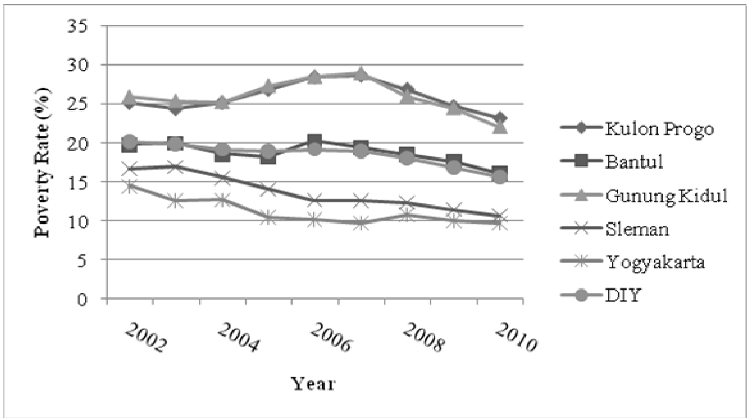
Poverty is a multidimensional problem. In macro, BPS uses basic needs approach. With this approach, poverty is seen as an economic inability to meet basic needs of food and nonfood measured from the expenditure. In other words, the poor are people who have an average per capita expenditure per month under a line called Poverty Line (GK). Number of poor people in DIY in 2002-2011 can be seen in Figure 1.

Figure 1 shows that the level of poverty in DIY is always higher than the national average. When viewed in more detail, the percentage of poor people in each district in DIY can be described as illustrated in Figure 2.



Source: BPS (www.bps.go.id)

Figure 1. The poverty rate of DIY and Indonesia in 2002-2011



Source: BPS (Daerah Dalam Angka)

Figure 2. Poverty rate of DIY and its district in 2002-2010

Figure 2 shows that in 2006 there was an increasing poverty rate in Kulonprogo, Gunung Kidul, and Bantul. This is occured because of the earthquake in DIY in 2006. Largest contributor to the poverty rate of DIY is Kulon Progo and Gunung Kidul. Thus, poverty reduction programs need to be focused on those two districts.

2. Parameter Estimated by GWR model

Brandon (2006) states that poverty is spatial problem. In analyzing spatial data, if spatial effects are ignored so the results of the analysis will be biased (Anselin and Griffith, 1988). Geographically Weighted Regression (GWR) is one of the most effective methods for estimating data that have spatial heterogeneity (Brunsdon, *et al*., 1996).

Steps to build GWR models are presented in Figure 3.

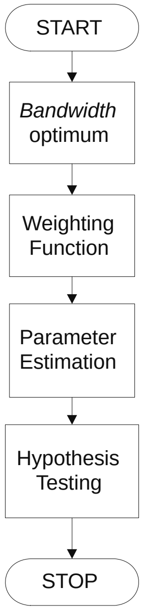
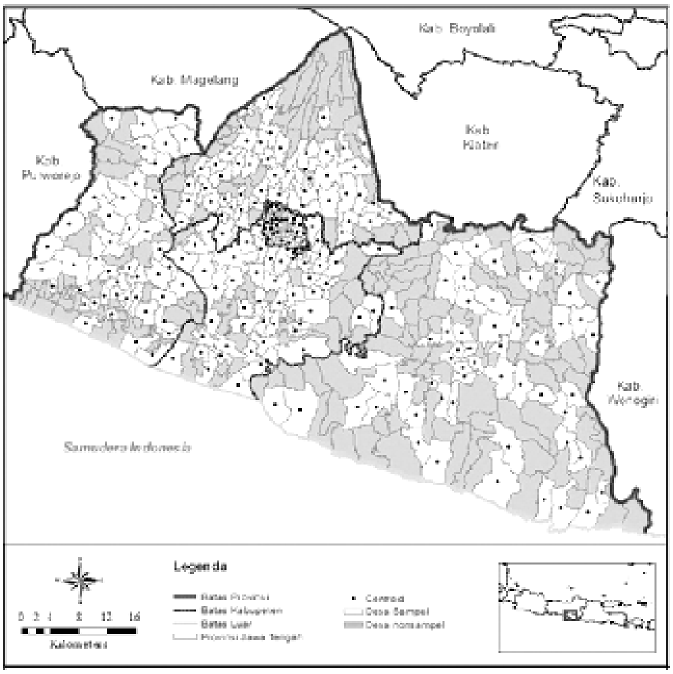


Figure 3. Steps to build GWR models



Source: BPS (Susenas 2011, data proccesing with ArcGIS 10)

Figure 4. Map of the distribution of Susenas 2011 samples

Determining the optimum bandwidth is based on the minimum AIC value 6579,80. This value is derived on the bandwidth 0,13 or 31 nearest villages from the observation. This study used adaptive bandwidth because the sampels of Susenas are not spread evenly (Figure 4). To obtain the weighting matrix, besides the bandwidth, Euclidean distance location (u_i, v_i) is also required for all sample locations. In this study, the Euclidean distance is measured from the center point between the villages with other villages. Equation (4) is used to obtain the estimated values of GWR parameters.

Table 2 shows the results of the parameter estimation of GWR and OLS models. GWR models will produce different parameter estimates at each location. Column 3 to column 7 consecutively shows descriptive statistics of the coefficients estimated by GWR models, namely the minimum, first quartile, median, third quartile, and maximum. Meanwhile, column 8 shows the value of the coefficient estimate OLS models is assumed to be constant at all locations.

Table 2 also providing information that GWR models can explain more phenomena than the OLS models that only produce one estimation parameter on each variable. For example, the relationship between population densities with an average per capita household expenditure can be positive or negative. Estimated value of the coefficient is positive, meaning that if there is an increase in the village population density, the average per capita household expenditure in the village will increase. It reflects that the population density increase in the village will increase welfare on that village. Meanwhile, the value of the coefficient estimated is negative which means that if there is population density increase in the village, the average per capita household expenditure in the village will decrease. It reflects the population density increase of the village will not increase the welfare.

Table 2. Parameter estimated of GWR and OLS models

No.	Variabel	Min	1^{st} Q	Median	^{3rd} Q	Max	OLS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	X_0	-7,25E+05	-5,64E+05	-4,55E+05	-1,81E+05	2,35E+06	-5,55E+05
2	X_1	-1,51E+01	-7,29E+00	2,26E+00	1,06E+01	2,53E+01	5,66E+00
3	X_2	1,21E+05	2,58E+05	3,44E+05	3,99E+05	4,82E+05	3,08E+05
4	X_3	-5,62E+04	-1,02E+04	-3,86E+03	-1,52E+03	2,53E+02	-2,92E+03
5	X_4	-8,44E+05	-4,30E+05	-9,60E+04	1,69E+05	1,33E+06	-4,76E+05
6	X_5	-2,62E+05	-1,16E+05	-7,08E+04	-3,79E+04	3,00E+04	-2,04E+04
7	X_6	-1,68E+06	-3,23E+05	-5,82E+04	1,26E+05	3,07E+05	-1,28E+04
8	X_7	-7,44E+04	-4,76E+04	-8,79E+03	7,22E+03	1,36E+05	-4,10E+04
9	X_8	7,66E+03	1,45E+04	1,86E+04	2,31E+04	2,96E+04	1,83E+04
10	X_9	-5,08E+03	2,50E+02	1,63E+03	1,85E+03	2,92E+03	1,00E+03
11	X_{10}	-6,21E+04	1,16E+05	1,52E+05	1,64E+05	2,26E+05	1,68E+05
12	X_{11}	2,50E+03	3,23E+03	3,45E+03	3,85E+03	4,72E+03	3,76E+03
13	X_{12}	-1,08E+04	-5,09E+03	-1,92E+03	1,67E+03	7,73E+03	-6,85E+03
14	X_{13}	-9,02E+04	-3,25E+04	-1,51E+04	-3,21E+03	6,63E+04	5,93E+03
15	X_{14}	-1,54E+05	-1,31E+04	-3,52E+03	3,36E+04	1,33E+05	2,73E+04
16	X_{15}	-7,06E+05	6,30E+04	1,24E+05	1,52E+05	1,94E+05	1,35E+05

Source: result from *software R 2.14.2*

These results support the research of Sumaryadi (1997) that population density may reflect different conditions of social welfare. Poor villages in urban areas are normally located in a sub-urban area with dense and irregular arrangement houses. Meanwhile, poor villages in the rural areas have low quality of human resources with sparse/less population density.

Coefficient estimates of GWR model for the variables of population density can be used to set the flow of population mobility. For example, village A populous and has negative coefficient estimated while village B sparsely populated and has a positive coefficient estimated. So, mobility can be directed from village A to village B, for example by opening new job vacancies in the area B. Thus, the development is expected to run optimally and evenly. Coefficient estimates of each village could be seen in Figure 6 (Appendix). In the meantime, in order to examine the relationship between response variable and predictor variables at each location required other studies that are not included in this study.

GWR ability to capture spatial nonstationarity should produce a better estimation than the global model. Some indicators can be used to compare global models and GWR model are smallest value of AIC, largest R^2, smallest $\hat{\sigma}$ and smallest Sum Square Residual (SSR).

From Table 3 it can be seen that based on all the best model criterion, GWR model are better than OLS model in estimating the average per capita household expenditure in DIY in 2011. Furthermore, to identify whether the GWR models can significantly explain the relationship response variable and predictor variables better than the OLS models, Analysis of varians (ANOVA) is used. The test results are presented in Table 4 as follows.

Table 3. Comparison of GWR and OLS models

No.	Kriteria	*GWR*	*OLS*
(1)	(2)	(3)	(4)
1	*AIC*	6579.795	6599,752
2	R^2	0,744872	0,5751
3	$\hat{\sigma}$	204200,9	231327,7
4	*SSR*	7,17E+12	1,19E+13

Source: Result from software R 2.14.2

In Table 4, there are GWR improvement component where its SSR is got from the difference between global model and GWR model. In other words, GWR improvement is a reduction due to the use of residual GWR models. By looking at the *p-value* in column 6 which is the approach of the *F* distribution with freedom degrees (194.056, 195.087), the null hypothesis can be rejected. In conclusion, with a 95% confidence level, GWR models provide more significant changes in explaining the relationship between the predictor variables and the response variable than OLS models. This means that the predictor variable with varies coefficient geographically able to explain the average per capita household expenditure in DIY better than the constant coefficients across geographic locations.

Table 4. Analysis of varians (ANOVA)

Source of Variation	*SS*	*Df*	*MS*	*F-stat*	*P-value*
(1)	(2)	(3)	(4)	(5)	(6)
OLS Residuals	1,1933e+13	16			
GWR Improvement	4,7673e+12	51,145	9,3212e+10		
GWR Residuals	7,1660e+12	171,855	4,1698e+10	2,2354	1,58e-08

Source: result from software R 2.14.2

3. Determination of Underdeveloped Villages

One objective of this study is to determine the status of a village. If the estimated value of the average per capita household expenditure in a village is under the poverty line, the area is classified as underdeveloped villages. According to BPS, the poverty line of DIY in 2011 was Rp 249.629,00. Based on the established modes from OLS and GWR, from 239 villages Susenas sample, the classification results as follows (Table 5).

Table 5 explains that the consistency of GWR model in classifying the villages is 91,63 percent, higher than the OLS model (89,96 percent). Thus, GWR models are more consistent in classifying the villages than OLS models.

To get the prediction of average per capita household expenditure on nonsampled areas, equation (12) is used. As a result, from 438 villages in DIY there are 13 villages belonging to the underdeveloped villages, presented in Table 6.

Table 5. Consistency of Classification (COC) OLS dan GWR model

Model	Classification	Original Data			*COC*
		Underdeveloped	Non-Underdeveloped	Total	(%)
(1)	(2)	(3)	(4)	(5)	(6)
OLS	Underdeveloped	1	6	7	89,96
	Non-underdeveloped	18	214	232	
	Total	19	220	239	
GWR	Underdeveloped	5	6	11	91,63
	Non-underdeveloped	14	214	228	
	Total	19	220	239	

Table 6. List of Underdeveloped villages by GWR Model

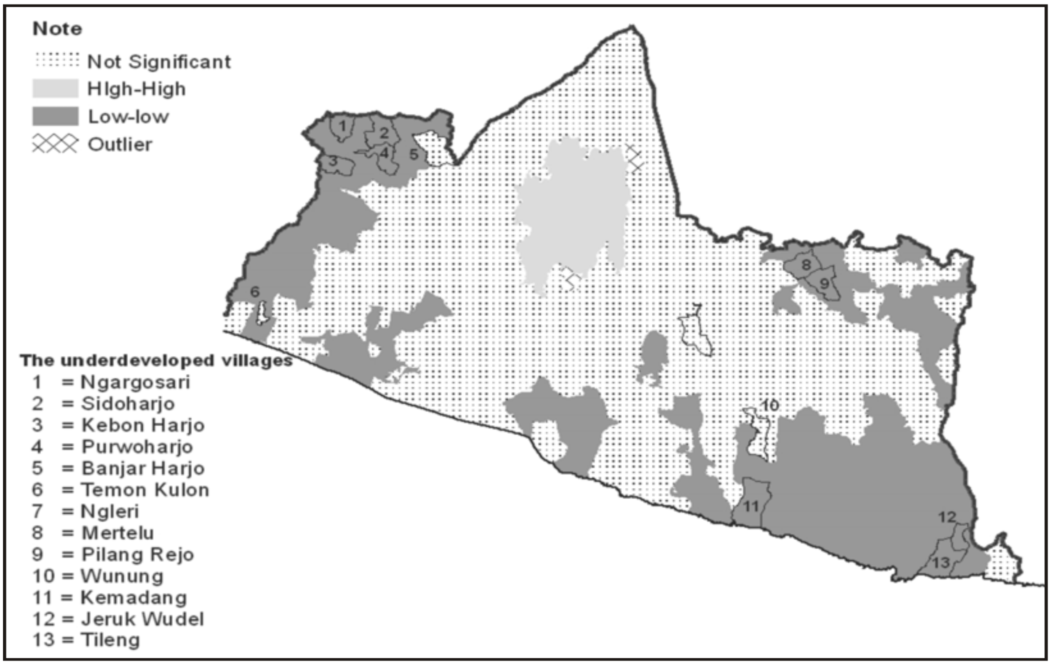
No.	District	Sub-District	Village
(1)	(2)	(3)	(4)
1	Kulon Progo	Temon	Temon Kulon
2	Kulon Progo	Kalibawang	Banjarharjo
3	Kulon Progo	Samigaluh	Kebonharjo
4	Kulon Progo	Samigaluh	Purwoharjo
5	Kulon Progo	Samigaluh	Sidoharjo
6	Kulon Progo	Samigaluh	Ngargosari
7	Gunung Kidul	Tanjungsari	Kemadang
8	Gunung Kidul	Girisubo	Tileng
9	Gunung Kidul	Girisubo	Jerukwudel
10	Gunung Kidul	Wonosari	Wunung
11	Gunung Kidul	Playen	Ngleri
12	Gunung Kidul	Gemdamg Sari	Mertelu
13	Gunung Kidul	Nglipar	Pilangrejo

4. Characteristics of underdeveloped villages

The poor society tends to clustering in certain places, forming proverty clusters in the region (Henniger and Snel, 2002). The existence of this cluster shows similarity or dissimilarity level of poverty in the neighboring area. Meanwhile, spatial autocorrelation can measure the strength of that spatial clustering (Anselin, 1995). Thus, detection of spatial autocorrelation can be used to identify the interaction between the region which causes a higher or lower concentration of poverty level. To see the spatial autocorrelation in underdeveloped villages in DIY, this study uses Local Indicators of Spatial Association (LISA).

Figure 5 is a map clustering tendency of the average per capita household expenditure estimated by GWR models. The blue, red, and yellow area are significant in a particular cluster indicating the interaction region. High cluster marked in red are in the city of Yogyakarta and the surrounding villages. It means there is a village which has high average per capita household expenditure which interacts due to high average per capita household expenditure in the surrounding villages. While the low clusters marked in blue are in the district of Kulonprogo, Bantul, and Gunung Kidul. It means there is a village which has low average per capita household expenditure which interacts due to lower average per capita household expenditure in the surrounding villages.

From Figure 5 it can be seen that there are nine underdeveloped villages in DIY was the low clusters significantly. Meanwhile, four other villages have low average per capita household expenditure, but the interaction was not significant.



Sources: Processing with Software Open GeoDa 10 and ArcGIS 10

Figure 5. LISA cluster map of the estimated average per capita household expenditure in DIY in 2011 by GWR models

Moreover, we get information about the condition of the 13 underdeveloped villages from Podes. Most of underdeveloped villages are lack in number of population (about 300 people per km square). They are located far from the office of regent, especially in Kulon Progo. Education facilities are good enough. There are a lot of medical facilities consist of Polindes and Posyandu, but they have no hospital. They also lack in the number of medical worker. The infrasturctures are not good enough. No one have public telephone and minimarket. About socio-economic factors, most people in underdeveloped villages are farmers. They use firewood to cook. Most of them subscribe electricity but they still depend on ground water, river, or rain to drink or cook. There are malnutrition problems in Kulon Progo. More detail information can be seen in Figure 7 (Appendix).

CONCLUSION

GWR models that noted spatial variation are able to explain the relationship between the response variable and the predictor variable better than the OLS models. By combining census and survey data, GWR models are very good to small area estimation. In this study GWR model are used to estimate the average per capita household expenditure in the village level. By using the GWR, we get result that there are 13 underdeveloped villages in the Province of Yogyakarta` in 2011, those are: Temon Kulon, Banjarharjo, Kebonharjo, Purwoharjo, Sidoharjo, Ngargosari, Kemadang, Tileng, Jerukwudel, Wunung, Ngleri, Mertelu, and Pilangrejo. Some of them are located in Kulonprogo and others are in Gunung Kidul.

Most of underdeveloped villages are located in Low-Low area. There are interaction between the village which have low expenditure and surrounding villages that also have low expenditure. Most of them are lack in number of population, public telecommunication facilities, minimarket, and medical worker. Most people in underdeveloped villages are farmers. They use firewood to cook. They also get water to drink and cook from ground, river, or rain

IMPLICATIONS

The results show that there are 13 underdeveloped villages in the Province of Yogyakarta in 2011. It can be a reference in determining the pro-poor development priorities and deliver direct assistance programs, such as the fuel subsidy.

To improve the welfare of people in underdeveloped villages, the government needs to develop the infrastructure in the field of communication. The minimal number of health is also a concern of the government that is expected to improve the health and tackling the problem of malnutrition. In addition, the government needs to encourage the banking sector to be more inclusive and better reach people in remote areas. By the access to capital, we hope that people does not depend on the agricultural sector so that the economy can develop more. PAM is necessary to expand because most of underdeveloped villages are still depend on ground, river, or rain.

Accelerating the development of underdeveloped villages, the government needs to trigger economic growth centers by the concept of 3D (Density, Distance, Division). Economic development in growth centers are expected to provide trickledown effect for the development of underdeveloped villages. First, the government took part in regulating the population density. Most of the underdeveloped villages is relatively low in population density so that the region lacked the human resources in the form of capital or labor. By increasing economic activity, such as construction and industrial markets, we hope the labours are interested to move to these locations. This is possibly occured due to increased economic activity will improve the productivity and quality of life. Besides, the government also needs to facilitate inter-regional mobility. Secondly, the government needs to reduce the distance by building good infrastructure to reduce transport costs. For example, by increasing the supply of cheap and convenient public transportation. Most of the underdeveloped villages in Kulonprogo located quite far away from the district capital (over 30 km), while Gunung Kidul is mountainous geographical location. Whereas, most of society work in the agricultural sector so that they find difficulties to sell the agricultural and buy needs that can not be found in the village. Third, the division relates to reduce inequalities between regions to create inclusive economic development. Another way in which the transition from the rural areas to educate people towards modern society, especially in matters of agriculture and the improvement of the wages system. In addition, managing the social gap can also be done through a fiscal transfer mechanism.

Good communication between the government and the people are absolutely necessary so that development can be optimized done. The development of underdeveloped villages is expected to push the poverty rate in Yogyakarta.

REFERENCES

Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L., 1992. *Spatial Data Analysis with GIS: An Introduction to Application in the Social Science*, *Technical Report of Nation Center for Geographic Information and Analysis*. Santa Barbara: University of California.

Anselin, L., 1995. Local Indicators for Spatial Association-LISA. *Geographical Analysis,* 27, 93 – 115.

Anselin and Griffith, 1988. Do Spatial Effects Really Matter in Regression Analysis? *Journal of the Regional Science Association,* 65, 11-34.

Badan Pusat Statistik, 2004. *Penyempurnaan Metodologi Perhitungan Penduduk Miskin dan Profil Kemiskinan.* Jakarta: Badan Pusat Statistik.

Badan Pusat Statistik, 2005. *Identifikasi dan Penentuan Desa Tertinggal 2002 Buku I: Sumatera*. Jakarta: Badan Pusat Statistik.

Badan Pusat Statistik, 2012. *Statistik Daerah Provinsi Daerah Istimewa Yogyakarta 2011*. Jakarta: Badan Pusat Statistik

Brandon, Vista., 2006. *Exploring The Spatial Patterns and Determinants of Poverty: The Case of Albay and Camarines Sur Provinces in Bicol Region, Philipines*, Thesis. Tsukuba: University of Tsukuba.

Brunsdon, C., Fotheringham A.S. and Charlton, M., 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28, 281-298.

Brunsdon, C., Fotheringham, A. S. and Charlton, M., 1999. Some Notes on Parametric Significance Tests for Geographically Weighted Regression. *Journal of Regional Science*, 39, 497-524.

Brunsdon, C., Fotheringham, A. S. and Charlton, M., 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: John Wiley and Sons, Ltd.

Dimulyo, Sarpono., 2009. Penggunaan Geographically Weighted Regression-Kriging untuk Klasifikasi Desa Tertinggal [The Using of Geographically Weighted Regression-Kriging for Underdeveloped Villages Classification]. In: *Seminar Nasional Aplikasi Teknologi Informasi 2009*. Yogyakarta: Universitas Islam Indonesia.

Ginting, Martinus., 1997. *Faktor-Faktor yang Berhubungan dengan Status Gizi Balita pada Empat Desa Tertinggal dan Tidak Tertinggal di Kabupaten Pontianak, Propinsi Kalimantan Barat Tahun 1995,* Tesis. Jakarta: Universitas Indonesia.

Henninger, N. and Snell, M. 2002. *Where Are The Poor? Experiences with the Development and Use of Poverty Maps*. Washington: World Resources Institute.

Hentschel, J., J.O. Lanjouw, P. Lanjouw, and J. Poggi, 2000. Combining Census and Survey Data to Trace the Spatial Dimension of Poverty: A Case Study of Ecuador. *The World Bank Economic Review*, 14.

Hurvich C M, Simonoff J S, Tsai C-L, 1998. Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion". *Journal of the Royal Statistical Society Series B,* 60, 271–293.

Leung, Y., Mei, C. L. and Zhang, W. X., 2000. Statistical Test for Spatial Nonstationarity Based on the Geographically Weighted Regression Model. *Journal Environment and Planning,* 32, 9–32.

Rahmawati, R. and Djuraidah, A., 2010. Regresi Terboboti Geografis dengan Pembobot Kernel Kuadrat Ganda untuk Data Kemiskinan di Kabupaten Jember [Geographical Weighted Regression with Bi-Square Kernel Weighting for Poverty Data in Jember]. *Forum Statistika dan Komputasi*, 15 (2).

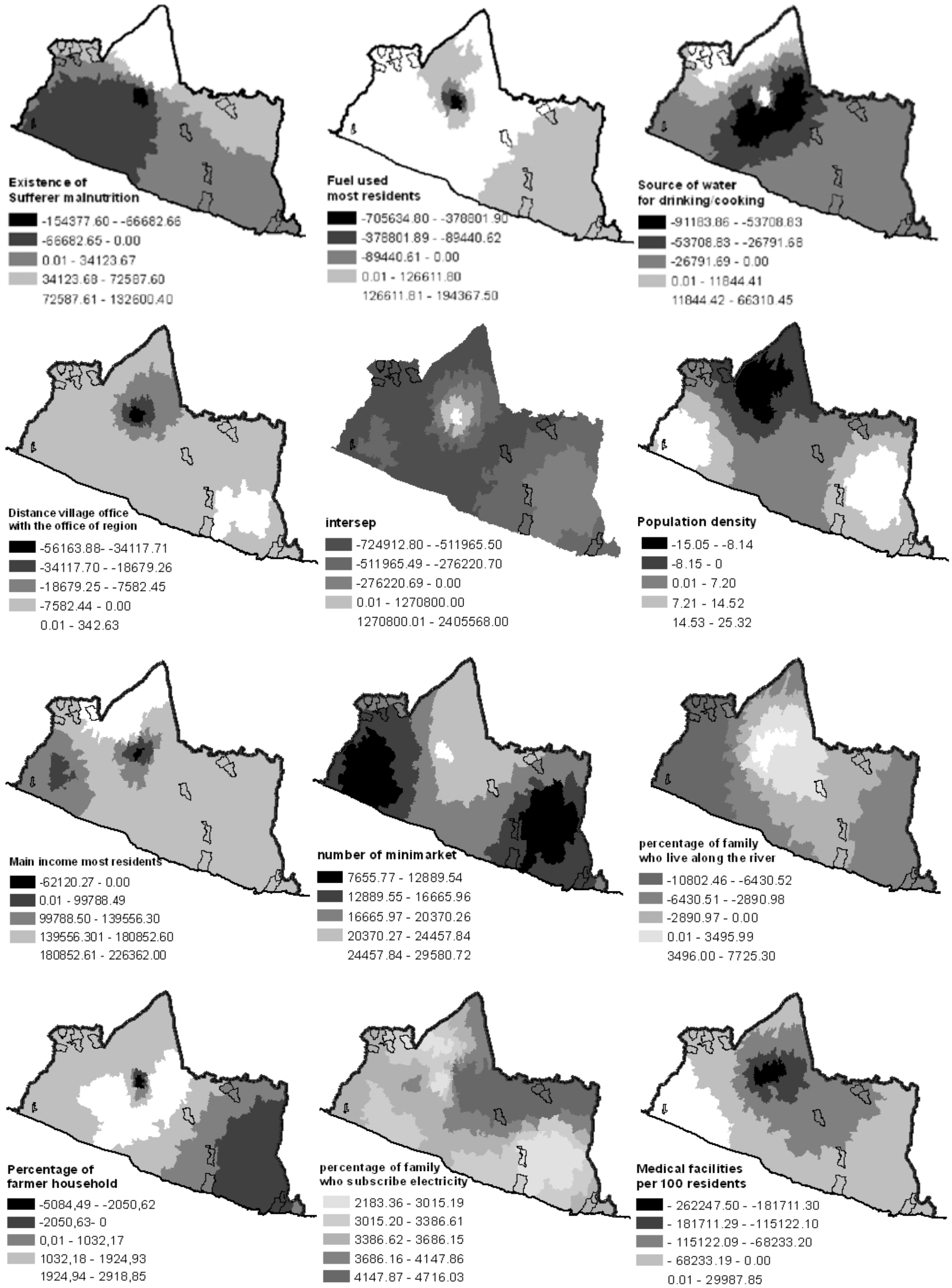
Sumaryadi, I Nyoman. 1997. *Pembangunan Masyarakat Desa dalam Perspektif Penanganan Kemiskinan Melalui Pelaksanaan Program Inpres Bantuan Pembangunan Desa dan Inpres Desa Tertinggal* [*Villagers Development in Poverty Handling through Implementation of Program Inpres Bantuan Pembangunan Desa and Inpres Desa Tertinggal Perspective*], Tesis. Jakarta: Universitas Indonesia.

Sukowati, Nisa N. *Pola Spasial dan Determinan Spasial dari Tingkat Kemiskinan di Nusa Tenggara Barat 2008* [*Spatial Pattern and Spatial Determinants of Poverty Level in Nusa Tenggara Barat 2008*], Skripsi. Jakarta: Sekolah Tinggi Ilmu Statistik.

World Bank, 2009. *World Economic Report 2009: Reshaping Economic Geography*. Washington DC: The World Bank.

Yrigoyen, C.C. and Rodriguez, I.G., 2008. Modelling Spatial Variations in Household Disposable Income with Geographically Weighted Regression. *Estadistica Espanola*, 50 (168), 321-360.

APPENDIX


Existence of
Sufferer malnutrition
-154377.60 - -66682.66
-66682.65 - 0.00
0.01 - 34123.67
34123.68 - 72587.60
72587.61 - 132600.40
Fuel used
most residents
-705634.80 - -378801.90
-378801.89 - -89440.62
-89440.61 - 0.00
0.01 - 126611.80
126611.81 - 194367.50
Source of water
for drinking/cooking
-91183.86 - -53708.83
-53708.83 - -26791.68
-26791.69 - 0.00
0.01 - 11844.41
11844.42 - 66310.45
Distance village office
with the office of region
-56163.88- -34117.71
-34117.70 - -18679.26
-18679.25 - -7582.45
-7582.44 - 0.00
0.01 - 342.63
intersep
-724912.80 - -511965.50
-511965.49 - -276220.70
-276220.69 - 0.00
0.01 - 1270800.00
1270800.01 - 2405568.00
Population density
-15.05 - -8.14
-8.15 - 0
0.01 - 7.20
7.21 - 14.52
14.53 - 25.32
Main income most residents
-62120.27 - 0.00
0.01 - 99788.49
99788.50 - 139556.30
139556.301 - 180852.60
180852.61 - 226362.00
number of minimarket
7655.77 - 12889.54
12889.55 - 16665.96
16665.97 - 20370.26
20370.27 - 24457.84
24457.84 - 29580.72
percentage of family
who live along the river
-10802.46 - -6430.52
-6430.51 - -2890.98
-2890.97 - 0.00
0.01 - 3495.99
3496.00 - 7725.30
Percentage of
farmer household
-5084,49 - -2050,62
-2050,63- 0
0,01 - 1032,17
1032,18 - 1924,93
1924,94 - 2918,85
percentage of family
who subscribe electricity
2183.36 - 3015.19
3015.20 - 3386.61
3386.62 - 3686.15
3686.16 - 4147.86
4147.87 - 4716.03
Medical facilities
per 100 residents
- 262247.50 - -181711.30
- 181711.29 - -115122.10
- 115122.09 - -68233.20
- 68233.19 - 0.00
0.01 - 29987.85

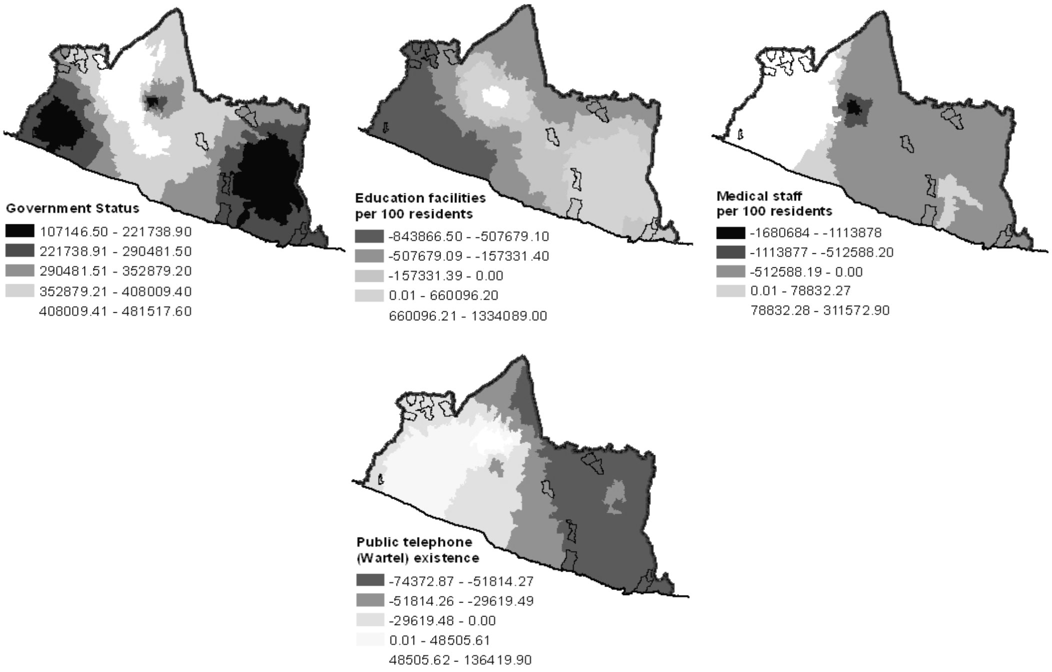
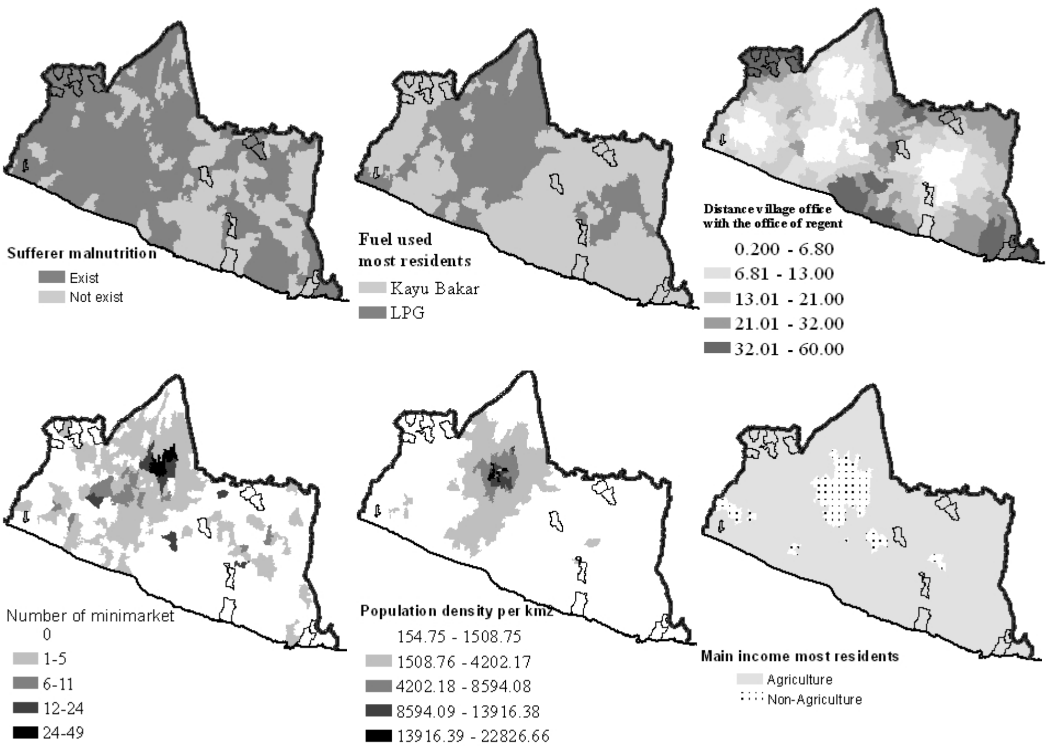


Figure 6. Map of the parameter estimated by GWR model



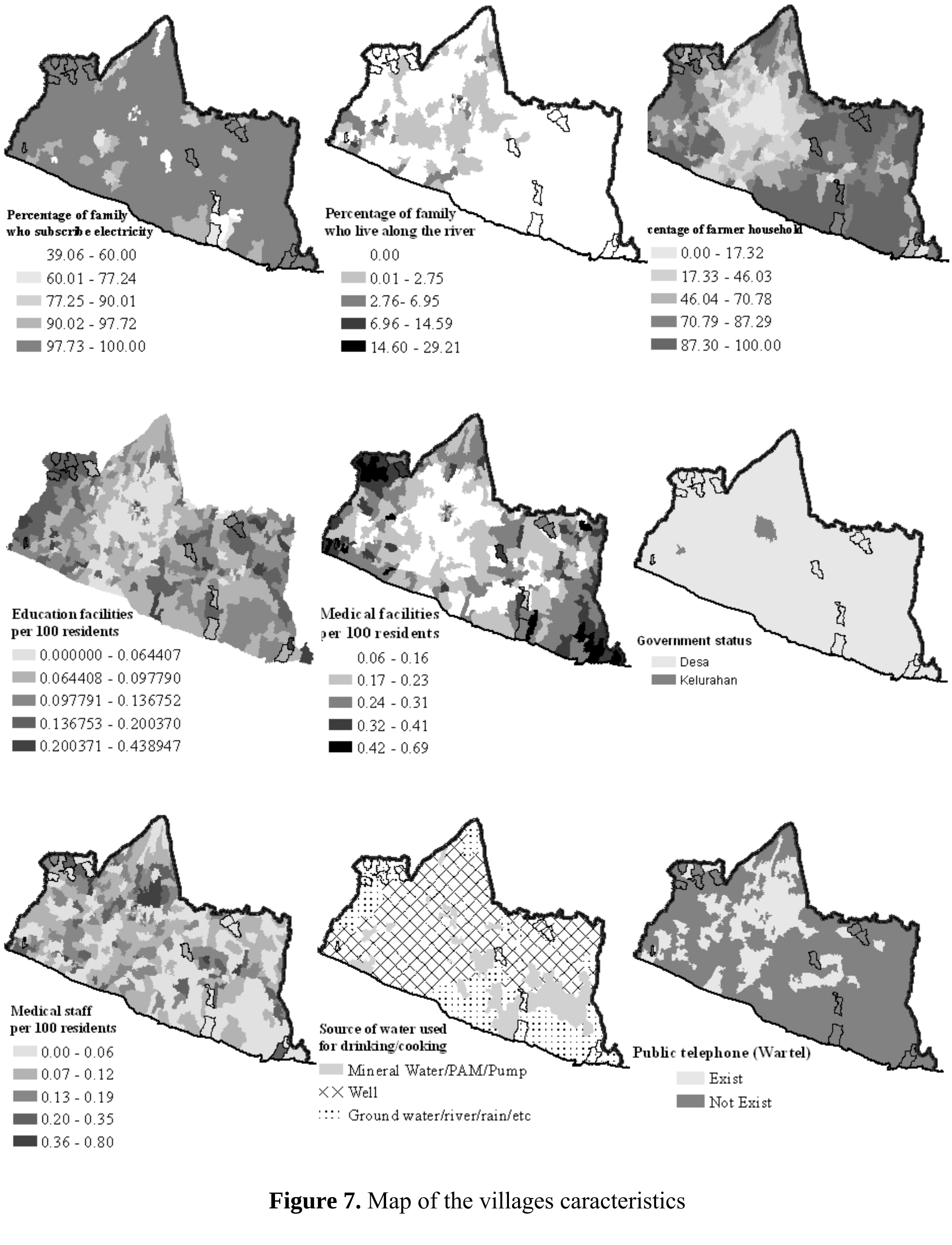


Figure 7. Map of the villages caracteristics